DNA replication and models for the origin of piRNAs

Jack R. Bateman[1] and Chao-ting Wu[2]*

Summary

The piRNA class of small RNAs are distinct from other small RNAs by their ~26–31 nucleotide size, single-strandedness and strand-specificity as well as by the clustered arrangement of their origins. Here, we highlight how these features are reminiscent of the mechanisms of DNA replication, and then present three models suggesting that the origin of piRNAs may be mechanistically similar to key processes in DNA replication. *BioEssays* 29:382–385, 2007. © 2007 Wiley Periodicals, Inc.

Introduction

Several recent reports have described a new class of mammalian small RNAs, called piRNAs, which are testis-specific, bound by the Piwi class of Argonaute proteins and distinct from siRNAs and miRNAs.[1–6] These piRNAs are single-stranded, primarily ~26–31 nucleotides (nts) in length, and largely encoded in ~100 to 200 loci scattered throughout the murine, rat and human genomes. Intriguingly, piRNAs derived from any single locus exhibit dramatic strand bias, in some cases numbering over a thousand with complementarity to only one strand of DNA. Furthermore, a substantial fraction of loci produce two divergent sets of piRNAs, where one set is complementary to one strand of DNA while the other is complementary to the other strand (Fig. 1A). The mechanism responsible for generatlng this novel class of small RNAs is as yet unknown; no evidence has been found for double-stranded or hairpin RNA precursors, which are observed for siRNAs and miRNAs, respectively. One model for the origin of piRNAs suggests that they arise from long transcripts that are then processed into smaller fragments.[1,4,6] However, as the ~26–31 nt length and single-strandedness of piRNAs are not characteristic of the 21–26 nt RNAs generated by Dicer-dependent mechanisms (reviewed by Ref. 7), the origin of piRNAs remains mysterious. Here we note that the size range of piRNAs in the vicinity of 30 nts, as well as their single-strandedness, their strand-specificity and the clustered nature of their genomic origins, are highly reminiscent of the mechanisms of DNA replication.

[1]Department of Genetics, Harvard Medical School, Boston MA.
[2]Division of Genetics and Division of Molecular Medicine, Harvard Medical School, Boston, MA.
*Correspondence to: Chao-ting Wu, Department of Genetics, Harvard Medical School, Boston MA 02115.
E-mail: twu@genetics.med.harvard.edu
DOI 10.1002/bies.20557
Published online in Wiley InterScience (www.interscience.wiley.com).

DNA replication

DNA replication typically proceeds in a bidirectional manner, involving two divergent leading strands of newly synthesized DNA and two complementary divergent lagging strands (reviewed by Ref. 8). Lagging strands grow via the sequential ligation of up to a thousand or more Okazaki fragments of ~100 to 200 nts, each of which carries an RNA–DNA primer at its 5′ end (Fig. 1B). The primers are initiated de novo by primase, which lays down ~8 to 12 RNA nts, after which DNA polymerase α (pol α) adds ~20 to 30 nts of DNA. When an RNA–DNA primer reaches the critical length of ~30 nts, replication factor C (RF-C) stalls pol α, causing pol α to be replaced by the processive DNA polymerase δ (pol δ). The Okazaki fragment then grows via the action of pol δ until it encounters and dislodges the 5′ end of the Okazaki fragment lying downstream. The dislodged 5′ flap, which can be as small as 2 to 3 nucleotides in size, is then removed by an interplay of nucleases such as flap endonuclease 1 and Dna2, and the resulting abutting Okazaki fragments are subsequently ligated together. Interestingly, in vitro studies reveal that flaps reaching a critical length of ~25–30 nts can become bound by replication protein A (RPA; also see Ref. 9) and released by Dna2. In contrast to the complexity of lagging strand synthesis, leading strand synthesis is simple; the RNA–DNA primer is extended without interruption, most likely by DNA polymerase ε.

Models for the generation of piRNAs

At the least, a juxtaposition of the arrangement of piRNAs with the structure of DNA replication forks highlights how the intrinsic asymmetry of DNA dictates corresponding asymmetries in processes for which it serves as a template. Might it also suggest piRNA synthesis and DNA replication to be mechanistically similar? In particular, both phenomena involve small strand-specific RNAs, and the divergent arrangement of piRNAs produced by some loci recalls the pairs of leading and lagging strands at origins of DNA replication. There are, however, arguments against a direct relationship between piRNA synthesis and DNA replication. For example, as piRNAs are generated from only a limited number of chromosomal sites and are not detected until pachytene, their synthesis cannot be a general feature of, and must be at least

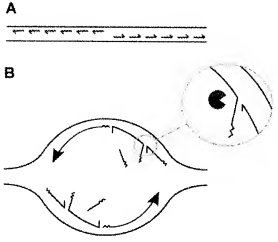

Figure 1. Origins of piRNAs and replication. **A:** Schematic of two divergent sets of piRNAs (red jagged lines) arrayed along the DNA strands (black lines) to which they are complementary. **B:** Schematic of bidirectional DNA replication. Leading strands have full arrowheads, while lagging strands and Okazaki fragments have one-sided arrowheads. RNA primers (red jagged lines) are typically ~8–12 nts long, while Okazaki fragments can be ~100–200 nts long in their entirety (not drawn to scale). Inset shows the 3′ end of one Okazaki fragment displacing the 5′ end of a downstream fragment, generating a flap that is subject to cleavage by a nuclease (blue pacman). Under certain circumstances, flaps can reach a length of ~30 nts, at which point they can become substrates for binding by RPA (not shown) and then cleavage by Dna2.

in part temporally offset from, whole genome S-phase DNA replication. Furthermore, the majority of piRNAs derive from loci that produce piRNAs of only one polarity and, in instances where a locus produces divergent sets of piRNAs, the relative 5′ to 3′ orientation of the two sets of piRNAs is divergent, while that of the RNA primers from a bidirectional pair of lagging strands is convergent (Fig. 1A,B). Finally, the RNA primers of Okazaki fragments are smaller than piRNAs and are thus unlikely to constitute piRNAs directly. Nevertheless, we wonder whether the mechanisms of DNA replication may warrant further consideration. Even if not directly responsible for piRNAs, they may shed light on the process by which these RNAs are produced.

First, the restriction of piRNA synthesis to just a few regions of the genome could be explained if the generation of piRNAs is associated with localized events, as observed for gene amplification (reviewed by Ref. 10), rather than with whole genome replication. In addition, DNA replication need not be bidirectional; termination can lead to polarized replication (reviewed by Ref. 11), which would be consistent with the unipolarity of piRNAs produced by some loci. With regard to the length of piRNAs, a primase-like RNA polymerase activity that does not yield to a DNA polymerase could produce RNAs that are significantly longer than the RNA primers of Okazaki fragments. In fact, a primase from the archaeon *Sulfolobus solfataricus* can extend RNA up to 1 kb in length (reviewed by Ref. 12). In the case of the leading strand, the resulting long RNA could be processed progressively from its 5′ end into a series of ~30 nt fragments via an RNA-specific endonuclease machinery analogous to that of RPA and Dna2 (Fig. 2A). Note that this model for generating ~30 nt strand-specific RNAs would be applicable to transcripts generated by any RNA polymerase, including RNA pol II. Alternatively, a polymerase that is interrupted by an RF-C-like factor when the oligomer it is producing reaches ~30 nt would also generate a series of ~30 nt strand-specific RNAs (Fig. 2B).

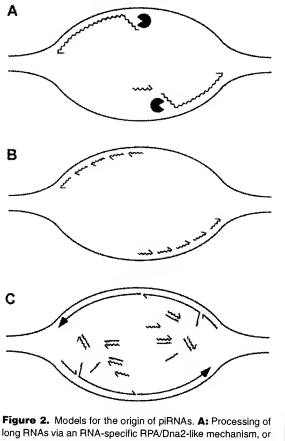

Figure 2. Models for the origin of piRNAs. **A:** Processing of long RNAs via an RNA-specific RPA/Dna2-like mechanism, or **B:** interruption of RNA synthesis by an RF-C-like factor, could produce strand-specific piRNAs. Both of these proposed mechanisms have the potential to also produce much longer RNAs, as precursors to piRNAs in the former case and as byproducts of noninterrupted RNA synthesis in the latter. **C:** Use of released ~30 nt flaps as templates for RNA synthesis could generate strand-specific piRNAs (not drawn to scale). In its simplest form, this lagging strand model does not predict the generation of long precursor or byproduct RNAs. Note that DNA and piRNA synthesis need not be concurrent (red; black, DNA).

A similar polymerase activity in lieu of lagging strand DNA synthesis could also produce ~30 nt RNAs. In the case of a bidirectional pair of lagging strands, however, this mechanism would generate two sets of RNAs whose relative orientation would be opposite to that of the divergent piRNAs arising from some loci (Fig. 1A,B). Rather, piRNAs could derive from the lagging strand through the action of an RNA polymerase on ~30 nt RNA–DNA fragments released by Dna2 or a Dna2-like nuclease during the previous round of DNA replication (Fig. 2C). Such polymerase activity is conceivable, as an RNA polymerase isolated from tomato leaves, although not yet from mammalian cells, can transcribe single-stranded RNA or DNA templates that are as short as 12 and 15 nts, respectively.[13] Note that this activity could provide a mechanism for the amplification of small RNAs in general. Importantly, constraining such a polymerase to initiate on DNA would correctly predict the strand-specificity observed for piRNAs.

Conclusions

In sum, we suggest that the size range, single-strandedness and strand-specificity of piRNAs as well as the clustered nature of their origins may indicate similarities between the mechanism by which they are generated and the process of DNA replication. The models proposed above could be addressed by determining whether piRNAs are associated with factors related to, or encoding functions similar to, those of DNA replication. In this light, it is intriguing that rat Piwi and piRNAs co-fractionate with the homologue to human RecQ1 DNA helicase,[4] which can be physically associated with, and stimulated by, RPA.[14] Importantly, a target bias for the cleavage or initiation site of an implicated nuclease or polymerase, respectively, could account for the tendency of piRNAs to initiate with uridine.[1–6]

Regardless of whether these ruminations reflect the true mechanism of piRNA synthesis, the discovery of piRNAs provides additional incentive to recall that every round of genome-wide DNA replication has the potential to generate millions of RNA and RNA–DNA fragments, which together may provide a distinct and potent cache of genetic material. These fragments could leave their sites of origin and, for example, alter gene expression or effect sequence changes. Indeed, a capacity for targeted mutagenesis has been well documented for synthetic RNA–DNA and DNA oligonucleotides, which have been proposed to act via their ligation to leading or lagging strands during DNA replication, or through their use as template or donor material during DNA repair (reviewed by Refs. 15–17). If RNA–DNA and RNA fragments generated during DNA replication were to perdure through organismal generations, they could even constitute the genetic cache hypothesized to underlie an unusual pattern of inheritance being investigated in mutant *hothead* strains of *Arabidopsis thaliana*[18–23] (however, also see Ref. 24). Ultimately, the encoding of genetic information in multiple forms should enhance the flexibility of developmental programs and mechanisms of inheritance, providing alternative strategies with which organisms can navigate their life cycles and evolution.

Acknowledgments

We would like to express our appreciation to Peter Burgers, Steve Elledge, Mitzi Kuroda, Nelson Lau, Marjori Matzke, Danesh Moazed, Charles Richardson, Gary Ruvkun, Anita Seto, and Johannes Walter for discussion and helpful comments.

References

1. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442:203–207.
2. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442:199–202.
3. Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 20:1709–1714.
4. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, et al. 2006. Characterization of the piRNA complex from rat testes. Science 313:363–367.
5. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, et al. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313:320–324.
6. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, et al. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev 20:1732–1743.
7. Matzke MA, Birchler JA. 2005. RNAi-mediated pathways in the nucleus. Nat Rev Gen 6:24–35.
8. Burgers PMJ, Seo Y-S. 2006. Eukaryotic DNA Replication Forks. In DNA Replication and Human Disease. Ed. by ML DePamphilis. Cold Spring Harbor Laboratory Press. p 105–120.
9. Fanning E, Klimovichm V, Nager AR. 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. Nuc Acids Res 34:4126–4137.
10. Tower J. 2004. Developmental gene amplification and origin replication. Ann Rev Genet 38:273–304.
11. Rothstein R, Michel B, Gangloff S. 2000. Replication fork pausing and recombination or "gimme a break". Genes Dev 14:1–10.
12. Lao-Sirieix S-h, Pellegrini L, Bell S. 2005. The promiscuous primase. Trends Genet 21:568–572.
13. Schiebel W, Hass B, Marinkovic S, Klanner A, Sänger HL. 1993. RNA-directed RNA polymerase from tomato leaves. J Biol Chem 268:11858–11867.
14. Cui S, Arosio D, Doherty KM, Brosh RM Jr, Falaschi A, et al. 2004. Analysis of the unwinding activity of the dimeric RECQ1 helicase in the presence of human replication protein A. Nuc Acids Res 32:2158–2170.
15. Court DL, Sawitzke JA, Thomason LC. 2002. Genetic engineering using homologous recombination. Ann Rev Genet 36:361–388.
16. Igoucheva O, Alexeev V, Yoon K. 2004. Oligonucleotide-directed mutagenesis and targeted gene correction: a mechanistic point of view. Curr Mol Med 4:445–463.
17. Parekh-Olmedo H, Ferrara L, Brachman E, Kmiec EB. 2005. Gene therapy progress and prospects: targeted gene repair. Gene Ther 12:639–646.
18. Lolle SJ, Victor JL, Young JM, Pruitt RE. 2005. Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis. Nature 434:505–509.

19. Chaudhury A. 2005. Plant genetics: hothead healer and extragenomic information. Nature 437:E1; discussion E2.
20. Comai L, Cartwright RA. 2005. A toxic mutator and selection alternative to the non-Mendelian RNA cache hypothesis for hothead reversion. Plant Cell 17:2856–2858.
21. Henikoff S. 2005. Rapid changes in plant genomes. Plant Cell 17:2852–2855.
22. Ray A. 2005. Plant genetics: RNA cache or genome trash? Nature 437:E1–2; discussion E2.
23. Krishnaswamy L, Peterson T. 2006. An Alternate Hypothesis to Explain the High Frequency of "Revertants" in Hothead Mutants in Arabidopsis. Plant Biol 9:30–31.
24. Peng P, Chan SW, Shah GA, Jacobsen SE. 2006. Plant genetics: increased outcrossing in hothead mutants. Nature 443:E8; discussion E8–9.